Welcome to Psy 652 Lab!
Module 9:
Multiple Regression

Objectives

Description of dataset

Multiple Regression vs. Simple Linear Regression (SLR)

Regression output with explanations

Variable descriptions of today's dataset (slp)

- The data set includes the following **predictor** and **outcome** variables:
  - **age:** Participant age
  - **anxiety**: Participant's level of general anxiety measured at the start of the study via a multi-item scale. The scale (average of all items) ranges from 1 to 7, where a higher score indicates a higher level of anxiety.
  - **hygiene**: Participant's sleep hygiene at week 6. It ranges from 0 to 10, and higher means better sleep practices.
  - **sleep**: Participant's average sleep efficiency during the month following the intervention, calculated as time spent in bed *asleep* (minus all the awakenings), divided by the total time spent in bed. It is expressed as a percentage.
  - **lifesat**: Participant's sense of life satisfaction measured 30 days after the completion of the intervention. It is a multi-item scale that ranges from 1 to 7, where a higher score indicates more satisfaction.

Differences between Multiple Regression and Simple Linear Regression (SLR)

- There isn't much!
- What changes our interpretation of the coefficients (slopes).
- Your R^2 and significance tests are interpreted the same!
- The only difference is that, now, we have to indicate that we are *controlling* for other variables in the model

Example regression results and interpretation (Outcome = Life satisfaction)

```
mod_ex <- lm(lifesat ~ age + anxiety, data = slp)
summary(mod_ex)
```

```
Call:
lm(formula = lifesat ~ age + anxiety, data = slp)

Residuals:
     Min       1Q   Median       3Q      Max
-2.26571 -0.56766 -0.02102  0.52970  2.22412

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.229341   0.181111  23.352   <2e-16 ***
age          0.025608   0.002552  10.035   <2e-16 ***
anxiety     -0.340978   0.036288  -9.396   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8029 on 597 degrees of freedom
Multiple R-squared:  0.2335,    Adjusted R-squared:  0.2309
F-statistic: 90.91 on 2 and 597 DF,  p-value: < 2.2e-16
```

Example regression results and interpretation (Outcome = Life satisfaction)

```
mod_ex <- lm(lifesat ~ age + anxiety, data = slp)
summary(mod_ex)
```

```
Call:
lm(formula = lifesat ~ age + anxiety, data = slp)

Residuals:
     Min       1Q   Median       3Q      Max
-2.26571 -0.56766 -0.02102  0.52970  2.22412

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.229341   0.181111  23.352   <2e-16 ***
age          0.025608   0.002552  10.035   <2e-16 ***
anxiety     -0.340978   0.036288  -9.396   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8029 on 597 degrees of freedom
Multiple R-squared:  0.2335,    Adjusted R-squared:  0.2309
F-statistic: 90.91 on 2 and 597 DF,  p-value: < 2.2e-16
```

All coefficients are significant

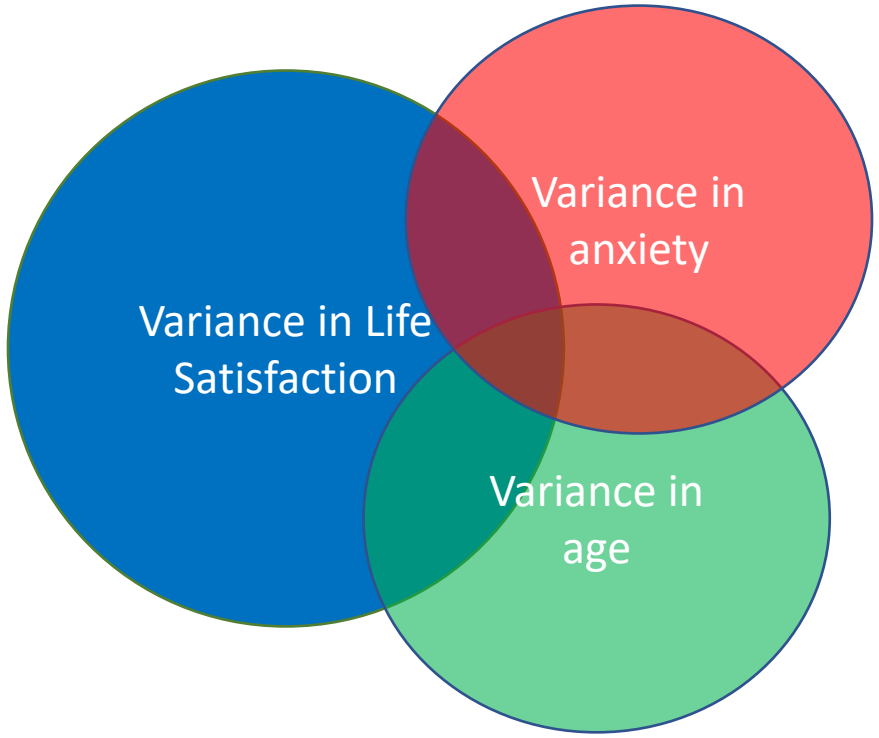
Our model explains 23.35% of the variance in Life Satisfaction

Overall model is significant

- (Intercept) = The expected life satisfaction score of someone ***at age 0 and an anxiety level of 0***
- age = For every one-unit increase in age, there is an expected .026 increase in life satisfaction ***WHILE CONTROLLING FOR ANXIETY.***
- anxiety = For every one-unit increase in anxiety level, there is an expected .034 decrease in life satisfaction ***WHILE CONTROLLING FOR AGE.***

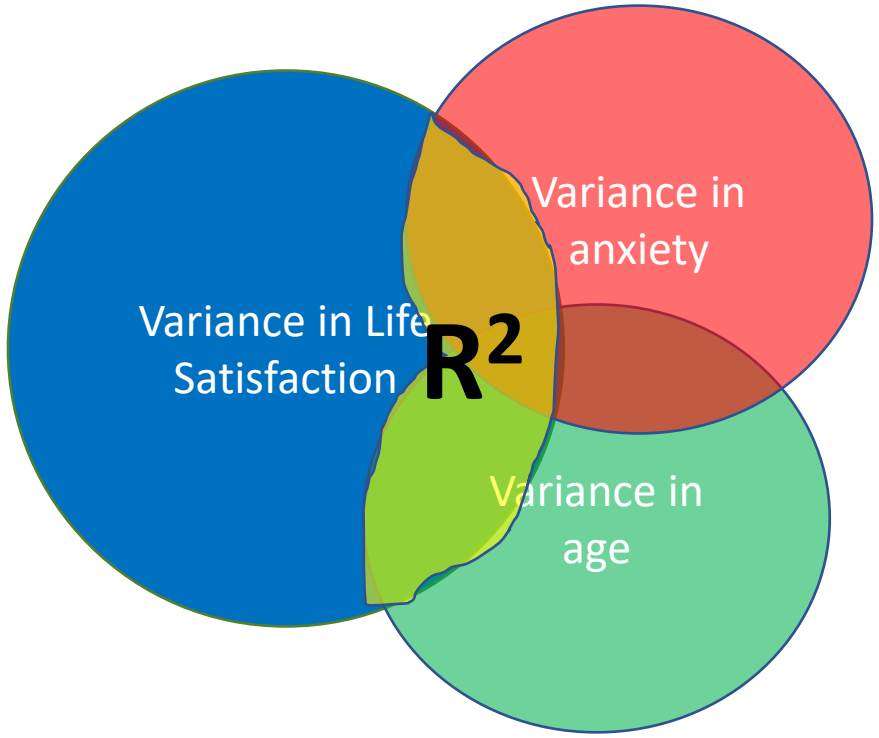
* Not drawn to scale

What is happening?



* Not drawn to scale

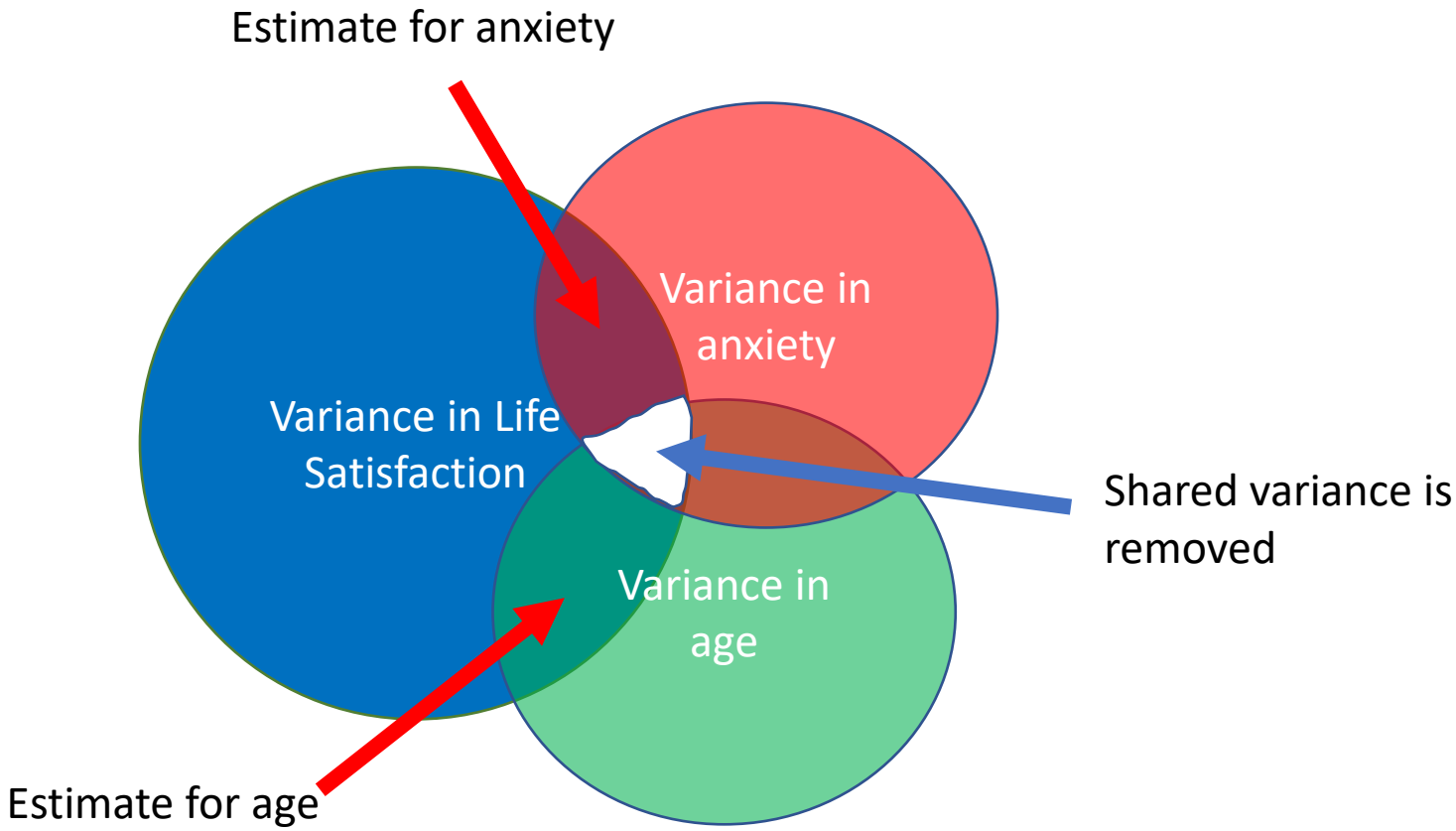
What is happening?: R^2



R^2 = The proportion of variance explained by the whole model (Even if it's shared)

* Not drawn to scale

What is happening?: Coefficients



Your coefficients are what is uniquely being explained by the predictor alone (excluding any shared variance).

Using the model to make predictions

```
mod_ex <- lm(lifesat ~ age + anxiety, data = slp)
summary(mod_ex)
```

```
Call:
lm(formula = lifesat ~ age + anxiety, data = slp)

Residuals:
     Min       1Q   Median       3Q      Max
-2.26571 -0.56766 -0.02102  0.52970  2.22412

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.229341   0.181111  23.352   <2e-16 ***
age          0.025608   0.002552  10.035   <2e-16 ***
anxiety     -0.340978   0.036288  -9.396   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8029 on 597 degrees of freedom
Multiple R-squared:  0.2335,    Adjusted R-squared:  0.2309
F-statistic: 90.91 on 2 and 597 DF,  p-value: < 2.2e-16
```

What is the expected life satisfaction of an individual that is age 52 and has an anxiety level of 5?

It's simple Algebra!

- $Y_{Lifesat} = \beta_0 + \beta_{age} + \beta_{anxiety}$
- Lifesat = Intercept + (.026 *Age) + (-.341*anxiety)
- Lifesat = 4.23 + (.026*52) +(-.341*5)
- **Lifesat = 3.877**

The expected life satisfaction of an individual that is age 52 and has an anxiety level of 5 is 3.877!

Let’s code!